한국정보전자통신기술학회논문지(jkiiect)'19-2, Vol.12 No.1
논문 19-12-01-22 http://dx.doi.org/10.17661/jkiiect.2019.12.1.22

Emergency Signal Detection based on Arm Gesture by Motion Vector Tracking in Face Area

Rabia Fayyaz*, Dae Jun Park* and Eun Joo Rhee**

Abstract This paper presents a method for detection of an emergency signal expressed by arm gestures based on motion segmentation and face area detection in the surveillance system. The important indicators of emergency can be arm gestures and voice. We define an emergency signal as the 'Help Me' arm gestures in a rectangle around the face. The 'Help Me' arm gestures are detected by tracking changes in the direction of the horizontal motion vectors of left and right arms. The experimental results show that the proposed method successfully detects 'Help Me' emergency signal for a single person and distinguishes it from other similar arm gestures such as hand waving for 'Bye' and stretching. The proposed method can be used effectively in situations where people can't speak, and there is a language or voice disability.

Key Words : Arm Gesture, Emergency Signal, Face Detection, Motion Vector, Tracking Points

1. INTRODUCTION

Gestures are the non-verbal way of communication, which are the body movements used to make actions to convey messages. Gestures include motions of the hand, arm, face, head and other parts of body, used in interactive games, remote control of home appliance, human-robot interaction, security systems and many other surveillance applications.

An emergency is a situation that brings an immediate risk to health, life or property. Arm gestures can be used as signals in an emergency situation such as violence, sexual abuse, car, and fire accident, etc.

Many methods have been developed to segment arm based gestures. Some researchers have used color information to segment hand images, and get arm information. Here, color is a variable factor that has failed to discriminate arm or hand with other color objects such as face, wall, etc. Additionally, color is affected by lighting and environmental conditions [1]. In other research, Kinect camera was used for structuring skeletons of the human body [2-3]. Here, the camera does not receive the infrared rays in outdoor environments because the reflected rays from the sun are more powerful than the infrared rays of the Kinect camera.

Thus, both methods [1-3] are useful only in indoor applications and failed in outdoor applications for the detection of arm gestures. Next, arm gestures detection is performed by Kinect camera for discriminating actions of eating and drinking using inertial body-worn sensors [4]. The method used motion and gesture segmentation with a sensor attached to the wrist and upper arm of the human

*Department of Computer Engineering, Hanbat National University
**Corresponding Author : Department of Computer Engineering, Hanbat National University (ejrhee@hanbat.ac.kr)
Received December 18, 2018 Revised December 25, 2018 Accepted January 24, 2019

body. Although this method is good, it needs a specific device.

Many other object segmentation methods are also available in literature such as background subtraction [5], optical flow [6] and frame difference [7], to segment hand, arm or other objects. The frame difference method uses the difference of two consecutive frames, and considers the objects to be moving if the difference is bigger than a threshold. Due to its simplicity and high performance in real time, this method is mostly used in research. Cho [8] used speech signals to detect the emergency but the method failed in the case of a person having speech disability and in the environment where person is unable to speak.

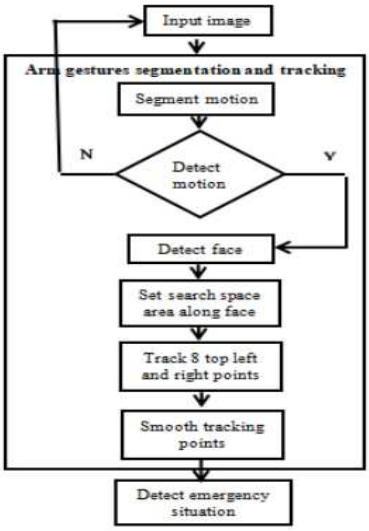

Fig. 1. The flow of the suggested method.

In this paper, we propose a method for detection of an emergency signal. The method is based on arm gestures detection without the use of color information. We use motion segmentation and face detection to recognize the arm gestures within a search area around the face. In order to recognize the specific arm gestures of interest, i.e. 'Help Me' arm gestures, we use N-tracking points within the search area of the face. Furthermore, we remove ambiguous tracking points by averaging the N tracking points to accurately maintain the start and end point of arm gestures in a noisy image. We trigger the emergency signal by counting frames when the horizontal direction of each motion has opposite signs for left and right arm gestures. The flow of the suggested method is given in Fig. 1.

The organization of this paper is as follows. The proposed algorithm is described in section 2. The usefulness of the proposed method is demonstrated through simulation in section 3. The conclusion is in section 4.

2. Arm Gestures for Emergency Situation

The 'Help Me' gesture is a commonly used emergency signal. It is the continuous simultaneous waving of the left and right arms, as shown in Fig. 2.

Fig. 2. Arm gestures in time series.
(a) At time 0. (b) At time 1. (c) At time 2. (d) At time 3. (e) At time 4. (f) At time 5.

It shows arm waving gestures in a series of time, where Fig. 2.(a), (b) and (c) show closing arm gestures at time 0, 1 and 2. Fig. 2.(d), (e) and (f) show opening arm gesture at time 3, 4 and 5. This closing and opening of arm waving gesture represents the 'Help Me' signal.

2.1 Arm gestures segmentation and tracking

First, we segment the arm gestures using motion and face information without any color information and additional hardware cost. The frame difference method detects motion in consecutive frames using the following equation (1), where f_{i-1} and f_i represent the previous and current frame to detect motion information (see Fig. 3.). This detected motion indicates that something has moved in the scene.

$$Diff = |f_i - f_{i-1}| \text{ --- (1)}$$

(a) (b) (c)

Fig. 3. Motion segmentation.
(a) Previous frame. (b) Current frame. (c) Motion segmentation.

Fig. 3. shows the motion segmentation in consecutive frames. Fig. 3.(a) and (b) show the current and previous frames, and Fig. 3.(c) shows the difference between the current and previous frames.

After the motion detection, the next step in the method is to perform face detection using Haar Cascade [9]. The reason for face detection along with motion detection is to consider the human arm gestures only.

And we create a search area around the face to segment arm gestures among other human actions. The search area box is much larger than the size of the face so that the method can capture more of the opening and closing 'Help Me' arm gestures.

After the arm gestures segmentation, the method tracks two N-point trajectories, beginning from top left and top right points respectively in the search area. Additionally, the average of every three points in each set of N-tracking points (left and right arm N-tracking points) is computed and used to overcome the ambiguity due to noise in the image (see Fig. 4.).

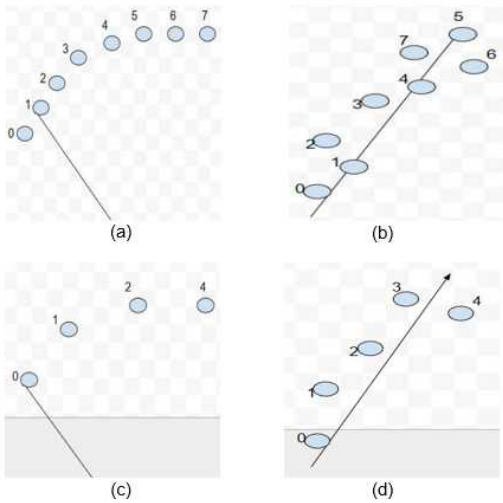

Fig. 4. Example of accurate and inaccurate arm-gesture. (a) Accurate arm-gesture of N-tracking points. (b) Inaccurate arm-gesture of N-tracking points. (c) Accurate arm-gesture of averaged N-tracking points of (a). (d) Accurate arm-gesture of averaged N-tracking points of (b).

Fig. 4.(a) and (b) show an example of N-tracking points with accurate and inaccurate

trajectories respectively. Fig. 4.(c) and (d) show the accurate trajectories after averaging N-tracking points, which overcomes the ambiguity, (see Fig. 4.(d)) and this accurately maintains the start and end point of the trajectories.

Fig. 5. shows the arm-gestures segmentation and tracking within the search area box. The small square and the big rectangle show the face and the search area respectively. Fig. 5.(a) shows the ambiguity of tracking points due to noise. Fig. 5.(b) shows averaged tracking points used to overcome the ambiguity and maintain the starting and ending point of arm gesture.

Fig. 5. Left and right arm gestures tracking within search area along face. (a) N-tracking points. (b) Average N-tracking points.

2.2 Detection of emergency situation based on arm gestures

Here, the detection of an emergency signal is elaborated based on repetitive and simultaneous waving of the left and right arms - 'Help Me' arm gestures - illustrated in Fig. 6. The detection process is performed using definition 1, which counts the horizontal opposite directions of left and right arm gestures. The method detects a 'Help Me' gesture when the count exceeds a specified threshold value, and it triggers an emergency signal.

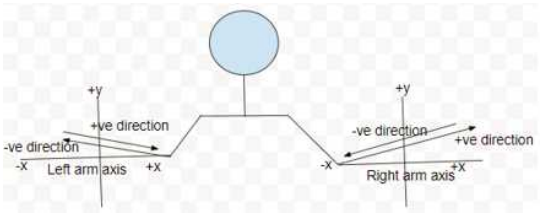

Fig. 6. Arm gestures for Help Me.

Fig. 6. shows 'Help Me' arm gestures which are the repetitive motion of the left and right arms in the positive and negative directions with respect to the horizontal axis.

We define the emergency signal in Definition 1. First, we take the difference of each consecutive pair of points (for example, L_{t0}, L_{t1} and L_{t2}, L_{t3} denoted by $\triangle L_0$ and $\triangle L_3$ respectively), and then we take the product of the differences, $\triangle L_x$. We do the same for the right arm to get $\triangle R_x$ and then increase count if both $\triangle L_x$. and $\triangle R_x$ are negative. When count exceeds the specified threshold, the alert is triggered.

[Def. 1] Gesture based Emergency signal

Given N-tracking points L_{t0}, L_{t1}, L_{t2}, L_{t3} for left and R_{t0}, R_{t1}, R_{t2}, R_{t3} for right arm, we define the emergency signal based on arm gestures as follows;

$$\text{if}(DL_x < 0 \;\&\; DR_x < 0)\ Then\ count\_help{+}{=}1$$

Where

$$DL_x = \Delta L_0 * \Delta L_3$$

$$DR_x = \Delta R_0 * \Delta R_3$$

$$\Delta L_0 = L_{t0} - L_{t1}$$

$$\Delta L_3 = L_{t2} - L_{t3}$$

$$\Delta R_0 = R_{t0} - R_{t1}$$

$$\Delta R_3 = R_{t2} - R_{t3}$$

$$\text{if (count\_Help} > threshold\text{)}$$

$$Then\ trigger\ emergency\ signal$$

$L_{t0}, L_{t1}, L_{t2}, L_{t3}$: x coordinates of the left arm
$R_{t0}, R_{t1}, R_{t2}, R_{t3}$: x coordinates of the right arm

3. EXPERIMENT AND DISCUSSION

We have performed the proposed emergency signal detection based on arm gestures to show its usefulness. This method was implemented in C++ and OpenCV [10] libraries using a CCTV camera.

We implemented the suggested method using real-time images taken from a CCTV camera. The real-time data included three different types of arm gestures; 60 gestures for 'Help Me', 15 for the 'Bye' and 30 for the 'stretching'.In the experiment of emergence signal detection, we confirmed the effectiveness of the proposed method which achieves a detection rate of 89.5%.

The reason why emergency signal is successfully detected is because our method uses the changes in the x coordinates of the left and right arms in a period of 1 or 2 seconds. So it is easy to distinguish the emergency signal from other gestures and other movements around.

Fig. 7. Gestures classified as "Help me".

Fig. 8. Gestures classified as "BYE or STRETCHING".

Fig. 7. shows arm gestures classified as 'Help Me' signal indicating an emergency, and Fig. 8. shows the other arm gestures. Each columns of Fig. 7. and 8. represent consecutive frames of specific gesture such as 'Help Me', 'Bye' and the 'Stretching'.

The 11% failure rate could be due to having more than one person waving their arms at the same time, arms moving out of the face area, the two arms not moving together, moving out of arms and the arms moving too slowly. Most failures are because they did not follow the definition of an emergency signal.

4. CONCLUSION

This paper described a method for detection

of emergency signals based on arm gestures in a surveillance system. The method determines 'Help Me' arm gestures based on motion and face detection. We use Haar-Cascade to detect the human face and frame difference for detecting arm gestures.

Arm gestures appear within a search area around the detected face. In this search area, the method tracks the repetitive arm gestures and it successfully recognizes the 'Help Me' gestures as an emergency signal. It also distinguishes the 'Help Me' gestures from similar arm gestures such as hand waving for Bye and body stretching for exercise.

The suggested system can easily detect an emergency signal made by a single person. And this method can be used effectively in situations where a person is not able to speak and people have language or voice disability. Additionally, it can complement sound and words such as 'Help Me' for detecting emergency.

In the future, we plan to extend our work of emergency signal detection for multiple human arm gestures in the scene.

REFERENCES

[1] M. Sigalas, H. Baltzakis and P. Trahanias, "Gesture recognition based on arm tracking for human-robot interaction", *Proc. of 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5424-5429, 2010.

[2] K. Lai, J. Konrad and P. Ishwar, "A gesture-driven computer interface using Kinect", *Proc. of 2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 185-188, 2012.

[3] K. Yamazaki, H. Miyao, and M. Maruyama. "Arm Gesture Recognition Using Continuous DP for User-Defined Gestures", *Proc. of HCI International 2013*, pp. 21-26, 2013.

[4] O. Amft, H. Junker, and G. Troster. "Detection of eating and drinking arm gestures using inertial body-worn sensors", *Proc. of* The *Ninth International Symposium on Wearable Computers*, pp. 160-163, 2005.

[5] S. Joudaki, M. S. B. Sunar, and H. Kolivand. "Background subtraction methods in video streams: a review", *Proc. of 2015 4th International Conference on Interactive Digital Media (ICIDM)*, pp. 1-6, 2015.

[6] L. A. Schwarz, A. Mkhitaryan, D. Mateus and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow", *Image and Vision Computing*, vol. 30, pp. 217-226, 2012.

[7] Z. Chaohui, X. Duan, S. Xu, Z. Song, and M. Luo, "An improved moving object detection algorithm based on frame difference and edge detection", *Proc. of Fourth International Conference on Image and Graphics (ICIG),* pp. 519-523, 2007.

[8] Y. I. Cho and S.S. Jang, "Implementation of intelligent Speech Recognition System according to CCTV Emergency Information", *Korean Institute of Intelligent Systems*, vol. 19, no. 1, pp 121-122, 2009.

[9] J. H. Yu and K. B. Sim, "Face Classification Using Cascade Facial detection and Convolution Neural Network", *Korean Institute of Intelligent Systems*, vol. 26, no. 1, pp 70-75, 2016.

[10] G, R. Bradski and V. Pisarevsky, "Intel's Computer Vision Library: Applications in Calibration, Stereo Segmentation, Tracking, Gesture, Face and Object Recognition", *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 796-797, 2000.

Author Biography

Rabia Fayyaz **[정회원]**

- Feb. 2013 : Hanbat National University, M.S.
- Feb. 2018 : Hanbat National University, Ph.D.
- Aug. 2018 ~ current : Hanbat National University, Artificial Intelligence and Computer Vision Lab., Researcher

〈Research Interests〉 Artificial Intelligence, Image Processing, Computer Vision, Pattern Recognition

Dae Jun Park **[정회원]**

- Aug. 2018 : Hanbat National University, B.S.
- Aug. 2018 ~ current : Hanbat National University, Artificial Intelligence and Computer Vision Lab., Researcher

〈Research Interests〉 Image Processing, Artificial Intelligence, Deep Learning

Eun Joo Rhee **[정회원]**

- Feb. 1989 : Chungnam National Univ., Ph.D.
- Aug. 1994.~ Aug. 1995 : Post-doctoral Fellow in Tokyo Institute of Technology
- Aug. 1989 ~ current : Hanbat Nat. Univ.., Dept. of Computer Eng., Professor

〈Research Interests〉 Artificial Intelligence, Image Processing, Computer Vision, Deep Learning